

## Reaching Out to Collaborators: Crowdsourcing for Pharmaceutical Research

Sean Ekins<sup>1,2,3,4,6</sup> and Antony J. Williams<sup>5</sup>

Received October 23, 2009; accepted January 5, 2010; published online January 27, 2010

Doing more with less resources used to be a situation common just for academic scientists. This is unfortunately still true for academics, but we are seeing others facing many of the same challenges. With the squeeze on budgets and cost-cutting resulting from recent worldwide economic challenges, the failure of many drugs to make it through the pipeline to the market, and the increasing costs associated with the drug development process, we are now seeing in the pharmaceutical industry a dramatic shift, perhaps belatedly, to have to accommodate similar challenges of doing more with less. This situation could also represent the further crumbling of a 150-year-old-plus paradigm of the large company being the predominant source for developing therapeutics for profit. We are also seeing increased discussion about different models of facilitating pharmaceutical research as well as the suggestion of opportunities to collaborate and use tools that perhaps would not have been considered in the past (1–4). This shift to “openness” in certain areas, specifically the sharing of pre-competitive data and processes, parallels the societal shifts we have seen in so many areas of open-source software development, the sharing of data, and the utility of free data resources and repositories, such as PubChem (<http://pubchem.ncbi.nlm.nih.gov/>) and others (see Table I). From the extreme of keeping entire projects in house, there is a shift to decentralized research. One view of pharmaceutical research is to use loose networks of external researchers from companies, academics or consultants, create a community around a shared interest and gather their ideas. We think this comes closest to the ideal of crowdsourcing where the wisdom of the many and their varied perspectives can benefit community-based

efforts whether in software, knowledge capture, etc. The loose definition of crowdsourcing as “outsourcing a task to a group or community of people in an open call” is a relatively new phenomenon, culture or movement, which is best summarized in the book “Wikinomics, How Mass Collaboration Changes Everything” (5). Living in our connected world, pharmaceutical researchers can communicate in a variety of ways (4) to leverage ideas from around the globe. These ideas do not have to come from within the walls of a single organisation. Taking this further: why limit access to just ideas? Open tools and data could feed an ecosystem. They could also breed a new class of researcher without affiliation, who has allegiance to neither company nor research organization. They test their hypotheses with data from elsewhere, they do their experiments through a network of collaborations, they may have no physical lab; while a shared cause may not be essential, confidentiality agreements and software may unite them as a loose cooperative. Such approaches may become more commonplace, like the Open Innovation efforts represented by companies such as NineSigma (<http://en.wikipedia.org/wiki/Ninesigma>) and Innocentive (<http://en.wikipedia.org/wiki/Innocentive>). The One Billion Minds approach for open innovation (<http://www.onebillionminds.com/>) has already been mapped into the Life Sciences, where a million minds in the community have been called to participate in community annotation in Wikiproteins (<http://genomebiology.com/2008/9/5/R89>).

A recent example of the power of crowdsourcing is the availability of freely accessible online resources to enable and support drug discovery. For instance, online databases, including PubChem, Chemical Entities of Biological Interest or ChEBI database (<http://www.ebi.ac.uk/chebi/>), DrugBank (<http://www.drugbank.ca/>), the Human Metabolome Database ([www.hmdb.ca](http://www.hmdb.ca)) and ChemSpider (<http://www.chemspider.com/>) represent good examples (6–8) in addition to commercial databases (9) and collaborative systems like CDD (<http://www.collaboratedrug.com>). These represent either government or privately funded initiatives with vastly differing resources and scopes. Chemistry (and with it biology) information on the internet has thus become more accessible just as we are seeing a massive increase in screening data coming from individual laboratories. Sometimes there are synergistic benefits of crowdsourcing; for example, the efforts behind the ChemSpider platform, originally a hobby project housed from a basement and recently acquired by the Royal

<sup>1</sup> Collaborations in Chemistry, 601 Runnymede Avenue, Jenkintown, Pennsylvania 19046, USA.

<sup>2</sup> Collaborative Drug Discovery, 1633 Bayshore Highway, Suite 342, Burlingame, California 94403, USA.

<sup>3</sup> Department of Pharmaceutical Sciences, University of Maryland, College Park, Maryland 21201, USA.

<sup>4</sup> Department of Pharmacology, Robert Wood Johnson Medical School, University of Medicine & Dentistry of New Jersey (UMDNJ), 675 Hoes lane, Piscataway, New Jersey 08854, USA.

<sup>5</sup> Royal Society of Chemistry, 904 Tamaras Circle, Wake Forest, North Carolina 27587, USA.

<sup>6</sup> To whom correspondence should be addressed. (e-mail: [ekinssean@yahoo.com](mailto:ekinssean@yahoo.com))

**Table I.** Examples of Crowdsourcing Resources for Pharmaceutical Research

Name	Website	Function
myExperiment	<a href="http://www.myexperiment.org/">http://www.myexperiment.org/</a>	Workflows, communities
DIYbio	<a href="http://diybio.org/">http://diybio.org/</a>	Community for do-it-yourself biologists
Protocol online	<a href="http://protocol-online.org/">http://protocol-online.org/</a>	Biology protocols
Open wetware	<a href="http://openwetware.org/wiki/Main_Page">http://openwetware.org/wiki/Main_Page</a>	Materials, protocols and resources
Open Notebook science challenge	<a href="http://onschallenge.wikispaces.com/">http://onschallenge.wikispaces.com/</a>	Crowdsourced science challenge—initially on solubility measurement
UsefulChem project	<a href="http://usefulchem.wikispaces.com/">http://usefulchem.wikispaces.com/</a>	Example of one scientist's open notebook
Laboratree	<a href="http://laboratree.org/pages/home">http://laboratree.org/pages/home</a>	Science networking site
Science Commons	<a href="http://sciencecommons.org/">http://sciencecommons.org/</a>	Strategies and tools for faster, efficient web-enabled scientific research
WikiPathways	<a href="http://www.wikipathways.org/index.php/WikiPathways">http://www.wikipathways.org/index.php/WikiPathways</a>	Curated biological pathways
Open Source Drug Discovery	<a href="http://www.osdd.net/home">http://www.osdd.net/home</a>	Collaboration around genomics and computational technologies

Society of Chemistry, has been acknowledged to have greatly enriched the content in the NIH's PubChem (9). We are also seeing crowdsourcing applied to get more perspectives on a problem, for example the annotation of 64 putative tools and probes from the NIH Roadmap MLSCN effort by scientists from different groups, using multiple filtering methods or molecule quality metrics (10).

What does the future hold for such databases and other crowdsourcing efforts, and what are some of the challenges to be aware of? While access to very large datasets as a starting point for biological information and modeling may be of value, there should be concerns regarding the quality of the compounds used for screening, e.g., will there be a high percentage of false positives? What about the fidelity of the data? Is the same batch of compound used by different groups? Are there experimental differences that result in large inter-lab differences in the manner in which they use technologies (11)? Do cell passage numbers differ? Are the internal standards the same? What is the diet of the animals used? What is the impact of dissolution variance (12)? Will the naïve user actually be able to dissect out the false positives or issues with data curation (13), which may represent a potential pressure point? What about issues with data protection, anonymity, ethics and tissue handling (14)? There are a myriad of other related questions and issues which could hamper merging data from different groups. On the opportunity side, there may be some obvious value in the smaller-scale experiments from individual laboratories being stored in a single location. Perhaps we can learn from the systems biology or network-building software community that have either manually or automatically annotated large databases (instances of object X interacts with object Y, either directly or indirectly) from individual experiments (15). Rarely is there kinetic data captured in these efforts, and yet if a database of such information could be created, this would become accessible. We can therefore see a need driven by the academic community predominantly for the curation of their single experiments in biology with benefits for preventing repetition and possible decrease of animal and reagent usage. This drive to curate biology can be encouraged by publishers and funding agencies, but once annotated in a desirable format (e.g. there would be a need to capture the

experimental protocol, and an ontology would be essential (16)) and location, the information could be freely available for other efforts, whether in data mining, SAR, software development or network building. The goal should be to bring scientists to a point where their data is shared and useful. It is one thing to provide large supplemental files with publications, but it is another to put the data in a location and format so that others can potentially learn from it. Perhaps *Pharmaceutical Research* (and for that matter other pharmaceutical journals) could play a role in ensuring that data within articles in the journal are deposited with freely accessible databases, such as PubChem or ChemSpider and beyond. Ultimately, we foresee there will be a highly networked structure linking the many crowdsourced database or other non-database tools to reduce redundancy. While we have already seen a dramatic growth in accessible databases, the innovation around computational methods for data analysis and mining have really not kept pace (13). There is an opportunity here for the scientific community to address these needs, and we may see a new wave of informatics company innovation. This could be catalyzed by public or private funding or even crowdsourcing X-prize type awards (<http://www.xprize.org/future-x-prizes/life-sciences>).

Perhaps there also needs to be some degree of focus initially to such an open drug discovery model to increase the probability of success, maybe around a neglected disease like Malaria or Tuberculosis (TB), or even rapidly emerging diseases (like swine flu), to demonstrate that it is more than a utopian concept. The incentive here could be that these diseases are rapidly becoming of more concern globally and could increase demand on healthcare resources (e.g. the reemergence of TB and ease of transmission). Targeted questions could be posed to the crowd regarding approaches to surmounting TB drug resistance, latency, target identification or developing novel delivery mechanisms (17,18). In addition, a gap analysis may be performed with the crowd to see what other novel issues may not have been considered.

As individuals, we are continually challenged by demands on our time and resources, both financial and intellectual, and participating in crowdsourcing neglected disease efforts would surely be a big motivating factor for many. Some

companies allow their employees to pursue personal projects as a small percentage of their time to foster creativity. Why not allow them to give back in this way and by contributing to open-source science, which may be another way to focus the research around their own areas of interest and skills? Perhaps governments can recognize the potential benefits and provide participating companies tax credits or other incentives. For example, the German government pays people to add to Wikipedia ([http://www.boingboing.net/2007/06/27/german\\_government\\_pa.html](http://www.boingboing.net/2007/06/27/german_government_pa.html)). With the stipulation of the Open Access policy by the NIH recently, government funds are effectively being directed in a manner that results in the release of data to the public very shortly after publication. This is an activity motivated by government grants.

For some in the “for profit” realm, the motive for much of their “open crowdsourcing” efforts is the revenue that is accrued from an innovation. For others, the motivation to participate in open drug discovery may not be financial but purely philanthropic in nature or simply the pursuit of an intellectual challenge. Think of it as the ultimate challenge where scientists collaborate with thousands of people to help global health. These two types of members of the crowdsourcing community could coexist. We would welcome suggestions from the various stakeholders with an interest in all aspects of the pharmaceutical R&D value chain on how open pharmaceutical collaborations could be facilitated. This is certainly an unsettling time in the industry, but after the storm has settled, we may be in a unique position to do further aspects of R&D differently and more cost-effectively, with implications for the whole scientific community and global healthcare. Less may indeed be more.

## ACKNOWLEDGMENTS

We are grateful to many discussions with colleagues on this topic as well as the reviewers suggestions.

**Conflicts of Interest** SE consults for Collaborative Drug Discovery, Inc. on a Bill and Melinda Gates Foundation Grant #49852 “Collaborative drug discovery for TB through a novel database of SAR data optimized to promote data archiving and sharing.” He is also on the advisory board for ChemSpider. AJW is employed by the Royal Society of Chemistry, which owns ChemSpider and associated technologies.

## REFERENCES

1. Bingham A, Ekins S. Competitive collaboration in the pharmaceutical and biotechnology industry. *Drug Discov Today*. 2009;14(23–24):1749–81.
2. Hunter AJ. The innovative medicines initiative: a pre-competitive initiative to enhance the biomedical science base of Europe to expedite the development of new medicines for patients. *Drug Discov Today*. 2008;13:371–3.
3. Barnes MR, Harland L, Foord SM, Hall MD, Dix I, Thomas S, *et al*. Lowering industry firewalls: pre-competitive informatics initiatives in drug discovery. *Nat Rev Drug Discov*. 2009;8: 701–8.
4. Bailey DS, Zanders ED. Drug discovery in the era of Facebook—new tools for scientific networking. *Drug Discov Today*. 2008; 13:863–8.
5. Tapscott D, Williams AJ. *Wikinomics: how mass collaboration changes everything*. New York: Portfolio; 2006.
6. Williams AJ. A perspective of publicly accessible/open-access chemistry databases. *Drug Discov Today*. 2008;13:495–501.
7. Williams AJ. Internet-based tools for communication and collaboration in chemistry. *Drug Discov Today*. 2008;13:502–6.
8. Louise-May S, Bunin B, Ekins S. Towards integrated web-based tools in drug discovery. *Drug Discovery—Touch Briefings*. 2009;6:17–21.
9. Southan C, Varkonyi P, Muresan S. Quantitative assessment of the expanding complementarity between public and commercial databases of bioactive compounds. *J Cheminformatics*. 2009;1:10.
10. Oprea TI, Bologa CG, Boyer S, Curpan RF, Glen RC, Hopkins AL, *et al*. A crowdsourcing evaluation of the NIH chemical probes. *Nat Chem Biol*. 2009;5:441–7.
11. Wang H, He X, Band M, Wilson C, Liu L. A study of inter-lab and inter-platform agreement of DNA microarray data. *BMC Genomics*. 2005;6:71.
12. Deng G, Ashley AJ, Brown WE, Eaton JW, Hauck WW, Kikwai LC, *et al*. The USP performance verification test. Part I: USP lot P prednisone tablets: quality attributes and experimental variables contributing to dissolution variance. *Pharm Res*. 2008; 25:1100–9.
13. Williams AJ, Tkachenko V, Lipinski C, Tropsha A, Ekins S. Free online resources enabling crowdsourced drug discovery. *Drug Discovery World*. 2010;11:1, Winter 2009/10.
14. van Veen EB. Obstacles to European research projects with data and tissue: solutions and further challenges. *Eur J Cancer*. 2008; 44:1438–50.
15. Ekins S. Systems-ADME/Tox: resources and network approaches. *J Pharmacol Toxicol Methods*. 2006;53:38–66.
16. Hoehndorf R, Bacher J, Backhaus M, Gregorio Jr SE, Loebe F, Pruffer K, *et al*. BOWiki: an ontology-based wiki for annotation of data and integration of knowledge in biology. *BMC bioinformatics*. 2009;10 Suppl 5:S5.
17. Balganesch TS, Alzari PM, Cole ST. Rising standards for tuberculosis drug development. *Trends Pharmacol Sci*. 2008;29:576–81.
18. Sacchettini JC, Rubin EJ, Freundlich JS. Drugs *versus* bugs: in pursuit of the persistent predator *Mycobacterium tuberculosis*. *Nature Reviews Microbiology*. 2008;6:41–52.